

# 비단조 뉴런에 의한 결정론적 볼츠만머신의 성능 개선

강형원, 박철영

대구대학교 정보통신공학부

## Performance Improvement of Deterministic Boltzmann Machine Based on Nonmonotonic Neuron

### Abstract

In this paper, We evaluate the learning ability of non-monotonic DBM(Deterministic Boltzmann Machine) network through numerical simulations. The simulation results show that the proposed system has higher performance than monotonic DBM network model. Non-monotonic DBM network also show an interesting result that network itself adjusts the number of hidden layer neurons. DBM network can be realized with fewer components than other neural network models. These results enhance the utilization of non-monotonic neurons in the large scale integration of neuro-chips.

### 1. 서론

신경회로망의 연구에서는 단순한 입출력 함수를 갖는 뉴런모델을 이용하여 네트워크를 다양하게 구축하여, 패턴인식이나 연상기억 등의 문제에 대해 많은 연구가 수행되어 왔다[1]-[3]. 여기에서 입출력 함수로서는 비선형 단조의 시그모이드(sigmoid)형 함수가 주로 이용되고 있다. 그러나 연상기억 문제와 같은 특정분야에 있어서 단조가 아닌 비단조 함수를 입출력 함수로 이용함으로써 기억용량이 증가되는 것이 모리타 등에 의해 보고되었다[4]. 여기서 이용한 함수는 입력값의 절대값이 어느 한계를 넘으면 출력값이 반전하는 시그모이드 함수로 Morita형 함수로 부른다. 이 결과는 그 후 여러 연구자들에 의해 수학적으로 해명되고 증명되었으며 뉴런의 입출력 함수에 비단조 함수를 이용하면 보다 고기능의 신경회로망이 실현될 수 있음을 의미한다[5]-[7]. 여기서 연상기억문제에서 다루는 신경회로망은 뉴런의 결합강도를 고정한 즉, 기억 매물형으로 학습에 의한 새로운 환경에 적응하는 것을 고려하지 않는다. 현재 학습기능을 갖는 신경회로망에 비단조 함수를 활성화 함수로 사용하는 뉴런을 이용한 연구 예는 거의 없다. 그러나 기억 매물형의 신경회로망에 있어서 비단조 뉴런의 중요한 효과는 기억용량의 증가이지만 학습 기능을 갖는 신경회로망에 대해서도 학습 성능의 향상이 기대된다.

본 연구에서는 학습기능을 갖는 결정론적 볼츠만 머신

에 비단조 뉴런을 이용하여 학습 성능을 수치 시뮬레이션을 통하여 분석한다. 먼저 네트워크의 은닉층에 비단조 및 단조뉴런을 이용한 경우에 대하여 각각 활성화 함수로 시그모이드 함수와 end-cut-off 타입의 비단조 함수를 사용한 경우에 대하여 성능을 비교한다. 실제 하드웨어로 구성하는 경우의 실현을 용이하기 하기 위하여 end-cut-off 타입의 비단조 함수는 연속함수가 아닌 구분선형 함수를 사용한다. 다음으로는 네트워크의 학습 성능을 학습수렴률을 이용하여 비교한다. 학습 대상으로는 N-Parity문제에 대하여 수행한다. 마지막으로 CMOS 회로를 사용하여 DBM 네트워크에 필요한 비단조 뉴런 회로를 설계하고 SPICE 시뮬레이션을 통하여 그 유용성을 입증한다.

### 2. 결정론적 볼츠만머신

볼츠만머신은 학습기능을 갖는 대표적인 신경회로망으로 대칭결합을 갖는 확률론적인 회로망에 학습규칙을 도입한 것으로 Hinton과 Sejnowski에 의해 제안되었다[8]. 특히 결정론적 볼츠만머신[9]은 확률론적 볼츠만머신에 있어서 뉴런의 출력에 평균장 근사를 이용하여 뉴런의 동작을 결정론적으로 하는 네트워크로서 Peterson과 Anderson에 의하여 제안되었다[10]. 평균장 근사에 의해 뉴런의 출력  $x_i$ 는

$$x_i = P(S_i=+1) \cdot (+1) + P(S_i=-1) \cdot (-1)$$

$$= \frac{1}{2} \left\{ 1 + \tanh \left( \frac{u_i}{T} \right) \right\} - \frac{1}{2} \left\{ 1 - \tanh \left( \frac{u_i}{T} \right) \right\}$$

$$= \tanh \left( \frac{u_i}{T} \right) \quad (1)$$

로 표현된다. 여기서 뉴런의 막전위  $u_i$ 는 다음 식으로 주어진다.

$$u_i = \tau \frac{du_i}{dt} + \sum_j w_{ij} x_j \quad (2)$$

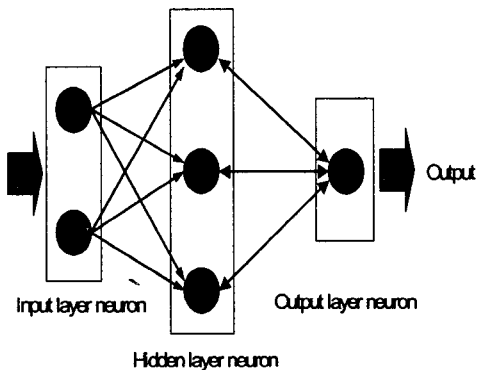
이때 상태  $a$ 에 있어서 학습에 의한 하중값 갱신량  $\Delta w_{ij}$ 는 다음 식으로 주어진다.

$$\Delta W_{ij} = \frac{\epsilon}{T} \sum_a \{ (x_i^a x_j^a)_{clamped} - (x_i^a x_j^a)_{unclamped} \} \quad (3)$$

여기서 clamped는 입출력 뉴런을 원하는 값으로 고정된 상태로 학습 phase로 부르며 unclamped는 입력 뉴런만을 고정된 상태로 반학습 phase라 한다. 식 (3)에 의해 하중값의 갱신량  $\Delta w_{ij}$ 는 각 상태에 있어서 학습 phase와 반학습 phase의 값  $x_i x_j$ 의 차를 모두 더한 값에 대해서 정수 ( $\epsilon/T$ ) 배이다.

### 3. 비단조 뉴런의 성능평가

네트워크는 <그림 1>과 같이 입력층 2뉴런, 은닉층 3 뉴런 그리고 출력층 1뉴런으로 구성되는 2-3-1네트워크이다. 출력층 뉴런은 단조 뉴런을 사용하여 입출력의 값은  $\pm 1$ 로 한다. 여기서 입력층 뉴런의 활성화 함수는  $y=x$ 의 단순 선형 뉴런이다.



<그림 1> DBM 네트워크 모델

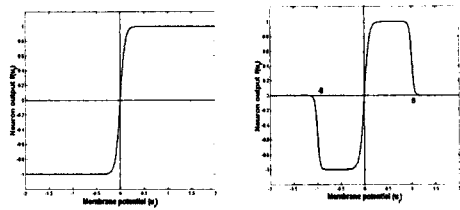
시뮬레이션에서 학습 수렴 조건은 <표 1>에 나타난 것과 같이 원하는 출력값과 네트워크의 출력값의 오차가 1% 미만이 되는 것으로 하고 최대 학습횟수는 2,000회

로 한다.

<표 1> 시뮬레이션 조건

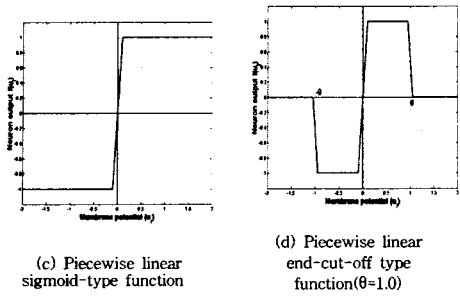
온도(T)	0.1(no annealing)
하중값 갱신방법	Batching learning
학습계수( $\epsilon$ )	0.004
초기하중값	[-0.01, 0.01]
최대학습횟수	2000회
목표값	{-1, 1}
샘플수	200

활성화 함수로는 일반적으로 <그림 2>의 (a)에 나타난 것과 같이 막전위의 값이 0을 넘으면 발화하는 시그모이드 단조함수이다. 한편 <그림 2>의 (b)는 end-cut-off 타입의 비단조함수를 나타내며 막전위의 절대값  $|u_i|$ 가 경계값  $\theta$ 를 넘으면 출력이 0이 되는 함수이다. 이 외에 막전위의 절대값  $|u_i|$ 가 경계값  $\theta$ 를 넘으면 출력값이 반전하는 Morita 타입 비단조 함수가 있다. 수치 시뮬레이션에서는 end-cut-off 함수를 이용하였다. 실제로 하드웨어로 구현하는 것을 고려할 때 <그림 2>의 (a), (b)와 같은 연속함수를 회로로 구현하는 것은 어렵기 때문에 하드웨어화를 위해서는 각각 <그림 2>의 (c) 및 (d)처럼 구분선형 함수를 이용한다. 이때 함수가 x축과 교차하는 부분의 경사의 절대치는 전부  $1/T$ 로 한다. 시뮬레이션에서는 <그림 2>에 나타난 4가지 함수를 이용하여 단조 뉴런을 이용한 경우와 비단조 뉴런을 이용한 경우의 성능을 비교하였다.

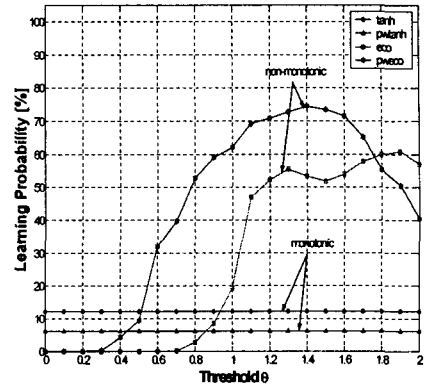


(a) Continuous sigmoid-type function

(b) Continuous end-cut-off type function( $\theta=1.0$ )



<그림 2> 네가지 타입의 활성화 함수

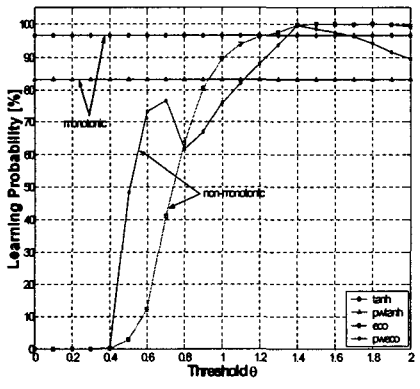


<그림 4> 4-4-1 네트워크에서의 4-패러티 문제에 대한 시뮬레이션 결과

4. 시뮬레이션

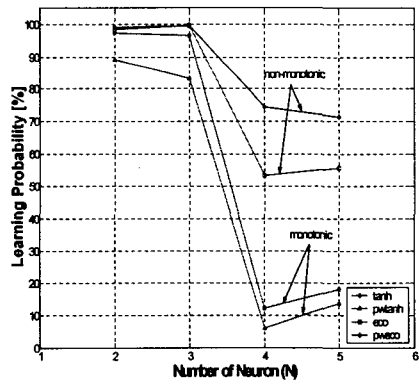
XOR는 입력이 두개인 특수한 경우의 패러티 문제라 할 수 있다. 따라서, 3-패러티 문제나 4-패러티 문제에서와 같은 비선형 수치 시뮬레이션에서도 단조와 비단조 사이의 학습 성능이 어떠한 차이를 두는지 비교하기 위하여 패러티 문제에 대한 시뮬레이션을 하였다.

<그림 4>을 보면 4-4-1 네트워크에서 단조함수의 경우 6%대의 학습 수렴률을 나타내고 있는 반면 비단조 함수의 경우 경계값  $\theta=1.6$ 에서 71%의 수렴률을 나타내고 있다. 이와 같이 입력 패턴이 증가하여도 비단조 함수는 단조 함수에 비하여 학습 수렴률 면에서 영향을 적게 받는 것을 알 수 있다.

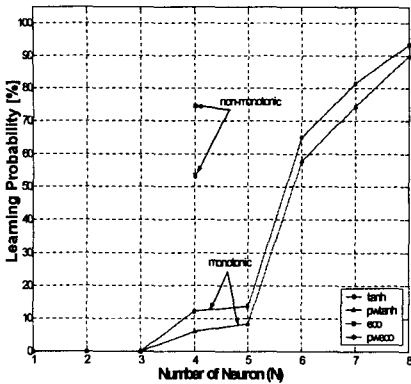


<그림 3> 3-3-1 네트워크에서의 3-패러티 문제에 대한 시뮬레이션 결과

<그림 3>에서 보면 3-3-1 네트워크에서 단조함수의 경우 높은 수렴률을 나타내고 있고, 비단조 함수의 경우 경계값  $\theta$ 를 조절함으로써 더욱 좋은 성능을 나타낸다. 예를 들면 단조 함수의 경우 96%의 수렴률을 나타내지만, 비단조 함수의 경우 경계값  $\theta=1.6$ 에서 약 100%의 수렴률을 나타내고 있다.



<그림 5> N-N-1 네트워크의 N-패러티 문제에 대한 시뮬레이션 결과



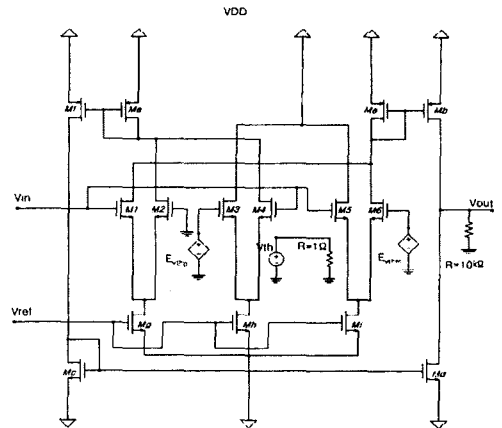
<그림 6> 4-X-1 단조 네트워크에서의 은닉층 노드수 변화에 대한 시뮬레이션 결과

<그림 5>은 N-N-1 네트워크에서 N의 갯수 즉, 입력층과 은닉층을 같이 증가시킬 때의 학습 성능에 대한 평가를 해 보았다. 우선, N의 갯수가 2개와 3개일 때에는 학습 성능 면에서 큰 차이가 없었으나 N의 갯수가 4개 이상으로 증가할 때 단조함수의 경우 낮은 학습 수렴률을 나타내는 것을 보였다. 이에 반해 비단조 함수의 경우 N의 갯수를 4개, 5개로 증가시켜도 학습 수렴률은 단조 함수에 비해 큰 차이가 없는 것으로 나타났다.

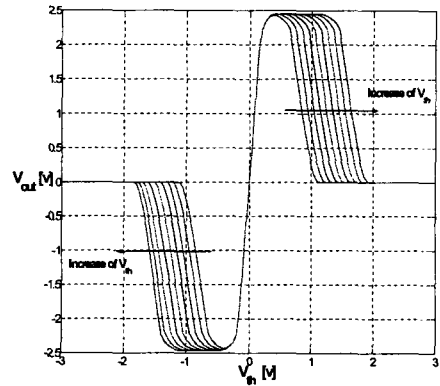
<그림 6>에서는 4-X-1의 단조 네트워크의 은닉층 수를 8개까지 증가시켜가면서 4-4-1 비단조 네트워크의 학습 수렴률과 동일해지는 은닉층 노드수를 찾아 보았다. 그림에서 보는 바와 같이 연속선형함수에서는 단조 함수의 은닉층 노드수가 6개 일때 비슷한 수치를 나타내었으며, 하드웨어 구현을 위해 사용한 구분선형 단조 함수의 경우는 은닉층의 노드수가 7개일때 비슷한 수치를 나타내었다. 그리고, 은닉층 노드수가 증가할수록 학습 수렴률이 높아지는 것을 알 수 있다.

### 5. end-cut-off 함수의 하드웨어 구현

본 연구에서 제시한 활성화 함수인 <그림 2> (b)의 end-cut-off형 함수를 전류미러(current mirror)를 활용하여 <그림 7>과 같이 회로를 구성할 수 있다. <그림 8>에는 그 시뮬레이션 결과를 보여준다. 시뮬레이션에 필요한 필수적인 파라미터를 <표 2>에 나타내었다.



<그림 7> CMOS 회로로 구성된 End-Cut-Off 함수



<그림 8> SPICE 시뮬레이션 결과

<그림 7>에서  $V_{in}$ 는 입력 전압의 값이며,  $V_{out}$ 은 end-cut-off형 함수의 출력을 나타낸다. 경계값  $\theta$ 에 해당하는  $V_{th}$  전압은 VCVS(Voltage Controlled Voltage Source)회로를 사용하여 조정하였으며,  $E_{vthp}$ 는 양의  $V_{th}$  전압을 나타내며  $E_{vthm}$ 은 음의  $V_{th}$  전압을 나타낸다. 또한, <그림 7>의 회로는 CMOS 트랜지스터의 차동쌍으로 표현되는데, 소자  $M_1$ 과  $M_2$ 는  $V_{in}=0$  일 때

$$f_{ecc}(u) = f_s(u) - \frac{1}{2}f_s(u - \theta) - \frac{1}{2}f_s(u + \theta) \quad (4)$$

의 첫 번째 항에서의 스텝 함수 역할을 한다.  $M_4$ 와  $M_6$ 는  $V_{in} = \pm V_{th}$ 일 때 식 (4)의 두 번째와 세 번째 항에서 스텝 함수의 역할을 한다.  $M_6$ 의 드레인 전류  $I_d$ 는  $M_1$ 과  $M_6$ 의 함으로 주어지고  $M_4$ 의 드레인 전류  $I_d$ 는  $M_2$ 와  $M_4$ 의 함으로 주어진다. 따라서 저항 R을 통

하여 흐르는 전류는  $I_b - I_d$ 로 주어지며 입·출력 전압 특성  $V_{out}(V_{in}) = R[I_b(V_{in}) - I_d(V_{in})]$ 에 의해 비단조 특성을 가지는 함수가 출력이 된다. 여기서,  $V_{out}$  단자의 저항 R은  $10k\Omega$ 로 하였다.

<그림 7>의 회로도 는 식 (4)을 하드웨어로 구성한 것으로서 앞에서 기술한 것과 같이 3개의 시그모이드 함수를 만드는 전류미러 회로를 이용한 복합적인 합성으로 end-cut-off 함수를 구현해낼 수가 있으며, 수치 시뮬레이션시 변경하여 실행한 경계값  $\theta$ 는 SPICE 시뮬레이션의 파라미터 값  $V_{th}$ 에 따라 자유로이 조정할 수 있다는 것을 알 수 있다.

<표 2> SPICE 시뮬레이션을 위한 파라미터 값

	$M_1 \sim M_6$	$M_a$	$M_b$	$M_c$	$M_d$	$M_e, M_f$	$M_g \sim M_i$
$W(\mu m)/L(\mu m)$	35/2	20/2	200/2	10/2	100/2	50/2	25/2
$V_{th}(V)$	1.0	-1.0	-1.0	1.0	1.0	-1.0	1.0
$K_p (\mu A/V^2)$	31	15.6	15.6	31	31	15.6	31

여기서  $V_{th}$ 는 zero-bias threshold voltage 이고  $K_p$ 는 transconductance parameter 이다.

<그림 8>에 나타난 시뮬레이션 결과처럼 CMOS 회로를 사용하여 본 논문에서 제시한 End-Cut-Off 회로를 구현할 수가 있으며, 시뮬레이션 파라미터 값과  $V_{th}$  값을 적당히 변화시킴으로서 DBM 네트워크에 필요한 비단조 뉴런에 관한 함수를 얻을 수 있다는 것을 알 수 있다.

## 6. 결론

본 연구에서는 학습기능을 갖는 대표적 신경회로망인 결정론적 볼츠만머신의 은닉층에 비단조 뉴런을 적용한 네트워크를 구성을 제안하고, 학습 대상으로 N-Parity문제에 관한 시뮬레이션을 수행하였다. 그 결과 단조 뉴런을 이용한 결정론적 볼츠만머신과 비교하여 end-cut-off 형을 이용한 비단조 뉴런은 경계값  $\theta$ 를 적절하게 조정함으로써 학습 수렴률이 향상 되는 것을 확인하였다. 이것은 비단조 함수가 복수의 단조 뉴런의 합성으로 구성되며 이에 기인하여 좋은 성능이 나타난다고 생각된다. 또한 패리티 문제에 대한 학습 성능의 평가 결과로부터 비단조 함수의 사용이 동일한 완성도를 유지하는 더욱 효과적인 하드웨어 구현이 가능하리라 생각된다. 향후 이들 결과에 근거하여 실시간 처리를 요구하는 제어나 고도 정보처리에 응용하는 방법에 관한 연구가 필요하다.

## 7. 참고 문헌

- [1] S. Nara, P. Davis, and H. Totsuji, "Memory search using complex dynamics in a recurrent neural network model", *Neural Networks*, vol.6, pp.963-973, 1993.
- [2] Y. Hayashi, "Oscillatory neural network and learning of continuously transformed patterns", *Neural Networks*, vol.7, pp.219-231, 1994.
- [3] C. Y. Park, Y. Hayakawa, K. Nakajima and Y. Sawada, "Limit cycles of one-dimensional neural networks with the cyclic connection matrix", *ICEICE Trans. on Fundamentals*, vol.E79-A, no.6, pp.752-757, 1996.
- [4] M. Morita, "Associative Memory with Nonmonotone Dynamics", *Neural Networks*, no.6, pp.115-126, 1993.
- [5] M. Morita, "Memory and Learning of Sequential Patterns by Nonmonotone Neural Networks", *Neural Networks*, vol.9, pp.1477-1489, 1996.
- [6] H. Yanai and S. Amari, "A Theory on a Neural Net with Nonmonotone Neurons", *Proc. IEEE Int. Conf. Neural Networks*, vol.3, pp.1385-1390, 1993.
- [7] Simon Haykin, 1999, *Neural Networks*, 2nd, Prentice Hall.
- [8] G. E. Hinton, and T. J. Sejnowski, "Learning and relearning in Boltzmann machines", *Parallel distributed processing*, vol. 1, pp. 282-317, 1986.
- [9] G. E. Hinton, "Deterministic Boltzmann Learning Performs Steepest Descent in Weight Space," *Neural Computation*, vol.1, pp.143-150, 1987.
- [10] C. Peterson and J. R. Anderson, "A Mean Field Theory Learning Algorithm for Neural Networks," *Complex Systems*, vol.1, pp.995-1019, 1987.
- [11] Tetsuya Asai, H. Yokotsuka, T. Fukai "A Mos Circuit for a Nonmonotonic Neural Network with Excellent Retrieval capabilities", *IEEE Trans. on Neural Networks*, vol.7, no.1, 1996.
- [12] C. A. Mead, 1987, *Analog VLSI and Neural systems*, Computer Science.